LF/MO3

LIMITED CIRCULATION

This paper is not an official document: it is for information only and must not be quoted in any printed publication.

Agricultural Memorandum No. 1054

Agromet Database II: Analysis and Design

by N.S. Callen and J. Portnall

SUMMARY

The analysis and design stages of the Agromet Database are discussed. Many well tried and trusted techniques based on the Structured Systems Analysis and Design Methodology (SSADM) approach were followed and one or two examples of these techniques are given.

1. Introduction

Designing a large database is a complex task and to assist the analyst/designer a methodical approach to the problem must be employed. For the Agromet Database an approach based on the Structured Systems Analysis and Design Methodology (SSADM) was adopted. SSADM provides a set of proven usable techniques and tools for designing computer-based information systems. Detailed rules and guidelines on when and how to use these techniques are included and documentation forms an integral part of the approach.

The two major phases involved in SSADM are System Analysis, defining what has to be done, and System Design, defining how it will be done. The objective of the Analysis study is to create a detailed specification of the User requirements and to agree with the Users what levels of service and performance are required. The objective of System Design is to define the structure and content of the database; how it is to provide the required services and how it is to be implemented.

The Analysis study for the Agromet Database commenced in early 1984. A few months later the design stage began and this was completed in Summer 1985.

2. Analysis Study

Following the guidelines set down by SSADM the analysis stage entailed identifying the flow of information, data and documents within the Agricultural Meteorology Section. Current programs used to produce routine agrometeorological output, routine and non-routine enquiries answered at HQ and outstations within the previous year, details of future agrometeorological products and further requirements of the database users were collected together and analysed. From them a list of all required meteorological data was drawn up ready for the design phase.

The user requirements of the new system were identified as follows:

i. The easy and efficient access to current agromet data especially with respect to location.

ii. The production of routine outputs without user intervention. In fact, all processes would be automated.

iii. The storage of historical agrometeorological data. This would allow the user to compare current data with data for previous seasons. Due to storage limitations it appeared that either one or two years data could be stored depending on the resources available. Both periods were acceptable to the users although two years of data was preferred.

iv. The storage of some data not currently available in an easily accessible form. These data include MORECS output and radar rainfall data.

v. The designation of space to be used as storage for future agromet products. This would allow new services to be implemented without the need for altering the database structure in any way.

vi. The improvement of data substitution procedures for missing data.

Having identified the data and user requirements of the new system the design of the database could then proceed.

3. Design Phase

The first task in the design phase was to analyse the required data and group each item under common headings such as MORECS square values, synoptic station values, etc.

Data items are the smallest units of named data in a database and the groups into which they are placed are known as records. A record is the basic and the smallest addressable unit in a database (ie any data item is accessed by searching through a record which itself may be addressed directly).

Secondly, all processes affecting the data throughout their storage life were identified. These processes included data extraction, quality control, substitution, derived data calculations and archiving amongst others.

Two major SSADM techniques were now employed concurrently to develop the database structure.

3.1 Data Flow Diagrams

This is a method of representing flows of information through a system and between the system and the outside world. Each of the processes identified earlier was divided into sub-processes and the records affected by each sub-process incorporated into a data flow diagram.

Figure 1 shows the 'daily data capture' process. Within the process boundary each sub-process is shown together with the temporary datasets used for local storage. The direction of data flow is indicated by an arrow.

Outside the process boundary the external entities and database records affected by the process are shown.

For reasons of clarity, the process has been simplified with references only made to synoptic stations. Consequently the numbering of the sub-processes does not appear to be sequential.

The detail shown in the diagram is much more than would be attempted in the initial stages of design. Only after accounting for user requirements and defining the logical data structure, discussed below, could the final data flow diagram be constructed.

3.2 Logical Data Structuring Technique

This technique represents a method for deriving entity models. An entity model is the data structure required to support the processing needs. When entity models are created, they are validated by ensuring that the required functions can be supported. If data structures cannot support the required functions, then they must be modified in order to do so.

Figure 2 gives the final Logical Data Structure of the Agromet Database. Records not only contain data groups, as defined earlier, but also necessary data accessing information such as the date, location details, etc.

Many relationships between records are identifiable within the structure. A relationship, defined a set, is schematically represented by a line with a crow's foot connecting two record types. Each set has one owner record and any number of member records where the member records are indicated by the crow's foot. For example, in the set 'Synoptic Station Details' (SSD), and 'Daily Synoptic Station Values' (DSSV), SSD is the owner record and for each of the station records there may be many DSSV records. Conversely, each DSSV record has only one owner SSD record within that set.

Records may be accessed either:

i. directly, indicated by an asterisk (*) or

ii. indirectly. A search is made through the member records of a set once the owner record has been accessed.

For consistency, any record type appearing on the Data Flow Diagram (see Figure 1) must appear in the Logical Data Structure.

3.3 Further Techniques

Other SSADM techniques employed were:

i. Entity Life History (ELH).

An ELH shows how an entity (ie record type) changes over time, what events the system needs to deal with, and in what order they occur. It provides the basis for keeping the system up-to-date, and illustrates the data as the basis for an operational system. Each event in the entity life history can be tied back to an entry on the data flow diagram.

ii. Process Outlines

This is a method of converting ELH's into definitions describing how transactions must be processed by the system. These definitions closely resemble program specifications to be followed during the implementation phase of the database.

The techniques used throughout the design phase compliment each other. Once the database design has been finalised the details acquired using these techniques must be consistent throughout.

3.4 Database Size

Computing resources obviously play an important role during the design phase of a database.

Table 1 gives a list of all record types defined using the above techniques. Also shown are:

i. The estimated number of occurrences of each record type.

ii. The length of each record type (bytes).

iii. Space usage for each record type (bytes).

The space usage is the product of the number of occurrences and the (record length + 8). Details of the space allocation within each record type are not given but allowances were made for the inclusion

of data pointers and indices required by the database management system. Many records also contain space for data flags which will be activated where substitution values are used or corrections have been made. These will enable the user to be aware of the exact status of the data.

The total size of the database is approximately 300 Mbytes (including the space necessary for temporary storage and program development etc). Except for the agrometeorological products and related records, radar rainfall data and some numerical forecast grid point data which are stored for 6 months, 7 days and one day respectively, all data will be stored for at least 24 months. Each month records which are more than 24 months old will be archived.

4. Further Development

Following the analysis and design phases the implementation of the database is the next phase. This commenced in Summer 1985 and entails much program development following the specifications laid down during the design phase.

Keyword: Database

Any comments or queries on this paper should be referred to the authors.

Meteorological Office Met O 3c
London Road
Bracknell
Berkshire RG12 2SZ

July 1985

Record Type	Estimated No. of Occurrences	Record Length (bytes)	Space Usage (bytes)
Radar rainfall day	7	20	196
Radar rainfall sq values	35,000	80	3,080,000
Radar rainfall sq details	5,000	52	300,000
Area/RR sq junction	6,000	28	216,000
Area	100	104	11,200
Grid pt. details	800	36	35,200
Grid pt. type	5	20	140
Product region	80	72	6,400
Daily products	39,040	828	32,637,440
Daily Weather Centre data	4,880	748	3,689,280
Weekly products	800	828	668,800
Area/Product region junction	300	28	10,800
Day	762	44	39,624
Synoptic station details	180	1,160	210,240
Daily synoptic stn values	137,160	864	119,603,520
Area/Synoptic stn junction	3,000	28	108,000
Rainfall station details	100	72	8,000
Daily rain station values	76,200	44	3,962,400
Area/Rain stn junction	500	32	20,000
Week	110	44	5,720
Weekly regional statistics	1,100	48	61,600
Weekly synoptic stn values	19,800	56	1,267,200
Weekly MORECS sq values	23,100	424	9,979,200
MORECS sq details	210	108	24,360
Area/MORECS sq junction	1,000	28	36,000
Month	25	36	1,100
Climat stn details	250	124	33,000
Daily climat stn values	190,250	96	19,786,000
Monthly climat stn values	5,500	36	242,000
Monthly MORECS sq values	5,250	404	2,163,000
Area/Climat stn junction	3,000	28	108,000
Daily information area	1	56	64
Fine mesh values 00Z	6	1,796	10,824
Fine mesh values 12Z	6	1,796	10,824
Coarse mesh values 00Z	6	376	2,304
Coarse mesh values 12Z	6	376	2,304
			198,351,880

TABLE 1: Sizing report for the Agromet Database Entities

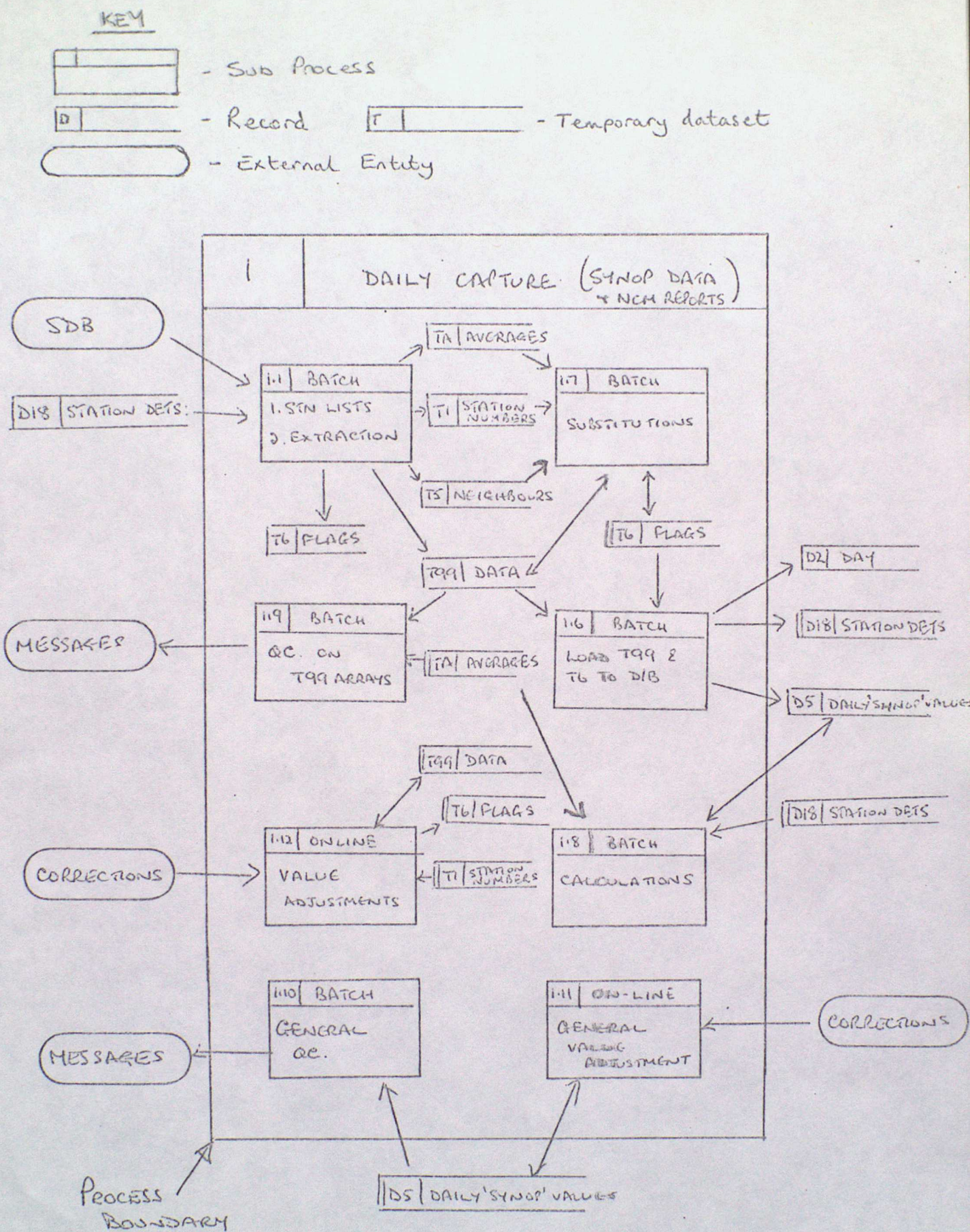

Fig. 1. Data Flow Diagram: Daily Data Capture Process

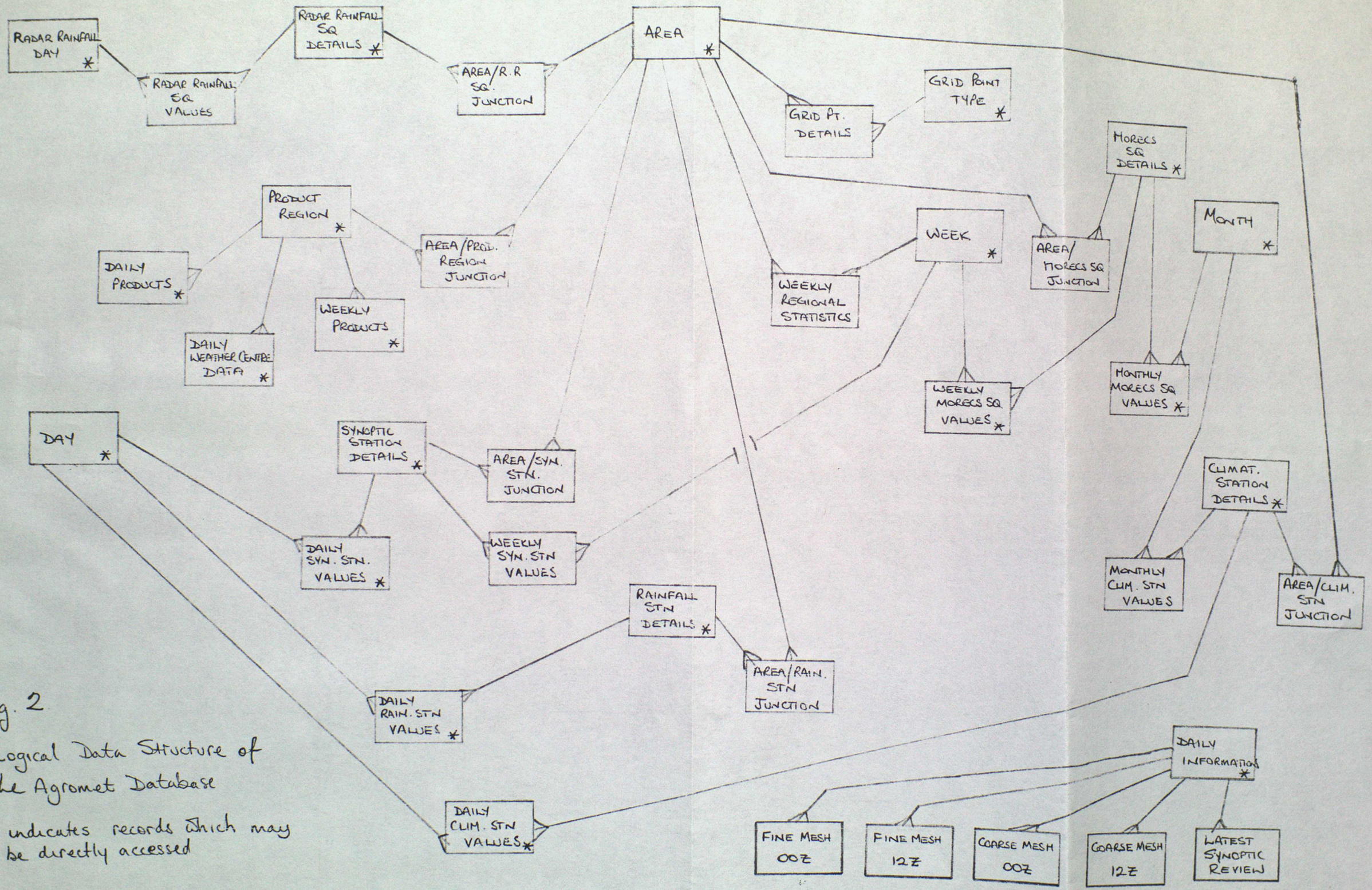

Fig. 2

Logical Data Structure of the Agromet Database

* indicates records which may be directly accessed